Problem statement - scoring compound v2 wallets.

Overview:

Zeru Finance is building an AI-powered, decentralized credit scoring system. You are provided with raw, transaction-level data from the Compound V2 protocol. Each record corresponds to a wallet interacting with the protocol through actions such as `deposit`, `borrow`, `repay`, `withdraw`, and `liquidation`.

Your task is to develop a machine learning model that assigns a **credit score between 0 and 100** to each wallet, based solely on historical transaction behavior. Higher scores indicate reliable and responsible usage; lower scores reflect risky, bot-like, or exploitative behavior.

Dataset:

You can access the dataset here:
Compound V2 Raw Dataset – Google Drive
(https://drive.google.com/drive/folders/1kCrMk30zlf8r1U4frgW9ecpYc1KhMSTE?usp=sharing)

Please **select any 3 files with the largest sizes** from the folder to ensure you're working with a significant portion of the protocol's activity.

Challenge Structure:

This is a self-driven, end-to-end modeling task. You are not provided with labels, predefined features, or a target column. Your responsibilities include:

- Defining criteria for "good" and "bad" wallet behavior
- Engineering features from raw transaction logs

- Choosing and justifying your modeling approach (e.g., clustering, rule-based scoring, supervised learning)
- Designing a credit scoring system that reflects behavioral quality and protocol health

Deliverables:

1. **Methodology Document**
   A concise, structured explanation of your scoring logic and rationale.

2. **Code Submission**
   A script or notebook that:
   - Loads and processes the raw data
   - Extracts wallet-level behavioral features
   - Outputs a score between 0 and 100 for each wallet

3. **CSV Output**
   A file containing scores for the top 1,000 wallets (sorted by score, highest to lowest).

4. **Wallet Analysis**
   A one-page document analyzing five high-scoring and five low-scoring wallets, explaining the observed patterns and their justification.

Constraints:

- Do not use any pretrained models, external labeled datasets, or third-party scoring systems.
- Do not use Zeru-provided code or schema definitions—derive everything independently from raw data.
- Your scoring logic must be non-trivial and irreproducible without your custom code and reasoning.

Evaluation Criteria:

- Clarity and originality of your methodology
- Quality and creativity in feature engineering
- Robustness, explainability, and consistency of the scoring model
- Insightfulness of wallet behavior analysis

For any questions or clarification, please reach out.